# Optimal Leniency Programs when Firms Have Cumulative and Asymmetric Evidence*

Marc Blatter† Winand Emons‡ Silvio Sticher§

Version: July 2014

**Abstract**

An antitrust authority deters collusion using fines and a leniency program. Unlike in most of the earlier literature, our firms have imperfect cumulative evidence of the collusion. That is, cartel conviction is not automatic if one firm reports: reporting makes conviction only more likely, the more so, the more firms report. Furthermore, the evidence is distributed asymmetrically among firms. Asymmetry of the evidence can increase the cost of deterrence if the high-evidence firm chooses to remain silent. Minimum-evidence standards may counteract this effect. Under a marker system only one firm reports; this may increase the cost of deterrence.

Keywords: antitrust, cartels, deterrence, leniency, evidence.

JEL: D43, K21, K42, L40

---

*We thank Andreas Bachmann, Juan Beccuti, Alain Egli, and Giancarlo Spagnolo for helpful comments. The usual disclaimer applies. The views expressed in this paper are those of the authors and do not necessarily represent those of the Swiss Competition Commission. Authors' address: Departement Volkswirtschaftslehre, Universität Bern, Schanzeneckstrasse 1, Postfach 8573, CH-3001 Bern, Switzerland.

†Universität Bern, Swiss Competition Commission, marc.blatter@vwi.unibe.ch.

‡Universität Bern, CEPR, CIRPÉE, Swiss Competition Commission, winand.emons@vwi.unibe.ch.

§Universität Bern, silvio.sticher@vwi.unibe.ch.

# 1 Introduction

A corporate leniency program reduces the sanctions for self-reporting cartel members. In 1993 the US Department of Justice revised its Corporate Leniency Program, committing itself to the lenient prosecution of the first confessor. It allows amnesty to be awarded even when an investigation has been started. This revision is considered as the most significant policy innovation in antitrust. It substantially increased the number of detected and convicted cartels.[1] The apparent success led the EU to adopt its own leniency program in 1996. Other countries followed suit.

With this paper we add to the burgeoning theoretical literature on leniency. This literature, which we review at the end of this section, has analyzed different effects of leniency. In this paper, we relax the following two assumptions typically made in this literature: First, if one firm reports the illegal behavior, the cartel is convicted for sure; each firm thus possesses perfect evidence. Inducing one firm to report, no matter which one, is sufficient for the antitrust authority (AA) to convict the cartel. Second, if each firm has perfect evidence about the illegal behavior, the distribution of the evidence is automatically symmetric.

By contrast, we look at the case where firms have imperfect and cumulative evidence. If a firm reveals its evidence, it increases the probability of conviction, but not necessarily to one. Put differently, in our framework the evidence of two firms leads to a higher probability of conviction than the evidence of only one firm.[2] Under imperfect evidence it is more difficult for the AA to deter collusion than under perfect evidence: the expected fine increases by less if a firm reports.

With imperfect evidence we can also allow for asymmetric evidence, i.e.,

[1] See, e.g., Harrington and Chang (2009) or Spagnolo (2008).

[2] Talking about our own experience at the Swiss Competition Commission (COMCO), evidence from whistle-blowers makes the COMCO's case stronger, but it is by no means taken for granted that the COMCO's ruling will be upheld on appeal.

one firm may have more evidence of the illegal behavior than another firm. We show that the asymmetry of the evidence may increase the AA's cost of deterrence. A high-evidence firm may be reluctant to apply for leniency even when a low-evidence firm does so. Finally, our framework allows us to analyze minimum-evidence standards as well as the effects of a marker system where the AA informs a firm whether leniency is still available.[3]

To be more specific, in our set-up two firms may repeatedly collude. If they do, they share the monopoly profit each period. Unilaterally deviating from collusion is attractive for the deviator and detrimental for the colluding firm. If the firms do not collude, they end up with competitive profits. The stage game is thus of the prisoners' dilemma type. Firms support collusion in the repeated game with grim-trigger strategies.

The AA wishes to deter collusion at minimal cost. The legislator has stipulated a fine for collusive behavior and a leniency program which grants full leniency to the first reporting firm. The AA chooses the probability that it starts an investigation as well as the effort it puts into the investigation. The AA's effort together with the evidence provided by the firms determines the probability of conviction.

In Section 3, we look at the scenario without a minimum standard of evidence and without a marker system. When a firm reports, it does not know its position in line, i.e., whether it gets leniency or not. For each possible level of the fine we characterize the AA's cost-minimizing enforcement policy. It turns out that the asymmetry of the evidence may increase the AA's enforcement cost as compared to symmetric evidence. With a high fine the AA needs little effort to deter collusion. With little effort by the AA, the evidence provided by firms is crucial. The high-evidence firm faces the following trade-off: If it reports, on the one hand, it has a chance to be granted leniency. On the other hand, it increases the probability of conviction sig-

[3]The only papers we are aware of analyzing imperfect and asymmetric evidence are Feess and Walzl (2005) and Silbye (2010). As to the best of our knowledge, we are the first to analyze a marker system formally.

nificantly. With little effort by the AA it therefore prefers to remain silent while the low-evidence firm reports.[4] The AA has to provide more effort to deter collusion as compared to the case where the evidence is symmetric and both firms report.

Within this framework we study the effects of a minimum standard of evidence in Section 4. If the standard is so low that both firms pass, it has no bite and our analysis without such a standard applies. A standard so high that no firm qualifies means effectively no leniency program and the AA does worse than without the standard. A standard such that the high-evidence firm qualifies while the low evidence firm does not lowers the AA's enforcement cost. Such a minimum standard provides the high-evidence firm with strong incentives to report: it can be sure to avoid the fine, the low-evidence firm cannot interfere.

In Section 5, we look at the case where the AA uses a marker system. After a firm applies for the marker, the AA informs the firm about its position in line. If the firm is first in line, it gets the marker and thus leniency if it chooses to report. The marker system allows firms to play the conditional strategy: report if the marker is available and do not report if the marker is no longer available. Firms play this strategy over a range of parameters where without the marker they would both report. When both firms play this conditional strategy, each of them is with equal probability first in line. This, in turn, means that the AA gets only one report, with equal probability the high or the low evidence. Thus, it gets less evidence than in the no-marker case where both firms report. More effort is necessary to deter this kind of collusion in the marker case than in the no-marker case.

Nevertheless, if firms are sufficiently asymmetric, the marker can be helpful for the AA. Without the marker, only the low-evidence firm reports while the high-evidence firm remains silent. With the marker, both firms play the

[4] This outcome, whereby a firm turns its fellow colluder in, runs contrary to the prediction of all previous models that either both or no firms apply; an exception is Harrington (2013).

conditional strategy and the AA gets the high or the low evidence with equal probability. Since it gets more evidence, it needs fewer resources to deter collusion than in the no-marker case. With minimum-evidence standards, the marker system is ineffective. If only the high-evidence firm qualifies for leniency, it can be sure to avoid the fine and a marker is not necessary for that privilege.

## 1.1 Related Literature

Our paper builds on the analysis of leniency programs by Motta and Polo (2003), Spagnolo (2003), Aubert et al. (2006), and Harrington (2008).[5] This literature analyzes the effects of leniency on the frequency of collusion. The main focus is to derive optimal fine structures; imperfect evidence is dealt with only in passing. In Aubert et al. (2006), for example, firms are able to partially destroy their evidence, thereby reducing the cartel's risk of being convicted. In Harrington (2008), the AA's stand-alone evidence varies over time, which affects the value of additional evidence provided by the firms. By contrast to our analysis, in all of the above-mentioned papers a cartel is convicted for sure as soon as one member submits evidence.

The issue of asymmetric and imperfect evidence was first addressed in Feess and Walzl (2005). Like in our framework, the firms possess different amounts of evidence about the illegal behavior. The AA optimally chooses a fine structure where the fine decreases with the amount of evidence provided. Silbye (2010) takes up this idea and extends the static Feess and Walzl set-up to repeated interactions. We touch on the issue of the optimal fine structure in Section 4. Nevertheless, following the institutional characteristics of the US leniency program, in our setting firms either get full or no leniency at all.

The focus of our analysis is the *cumulative* nature of evidence. Both

[5]Further recent theoretical research includes Harrington and Chang (2009), Harrington (2013), and Sauvagnat (2014); empirical and experimental research includes Bigoni et al. (2010), Brenner (2009), and Miller (2009). For surveys of the earlier literature, see Rey (2003) and Spagnolo (2008).

Feess and Walzl (2005) and Silbye (2010) assume that only one firm applies for leniency. The AA cannot get hold of the non-reporting firm's evidence. Thus, they implicitly assume that a firm observes if the other firm reports. We deal with this informational issue explicitly. Under a leniency program with a marker system a firm finds out whether its conspirator applied for leniency. In the absence of a marker system, a firm does not know whether the other firm applied for leniency. It is possible that the AA benefits from having more than one leniency application because more evidence strengthens its case.[6]

## 1.2 Institutional Background

In both the US and in the EU, the institutional details of a leniency program are not determined by competition law, but rather by internal documents of the AAs.[7] This implies that the AA can revise its leniency program as to what it sees fit. Therefore, in our analysis, we only specify a limited number of details and allow the AA to choose between several institutional designs.

Typical features of leniency programs include: first, the assignment of leniency to the first applicant who reports participation in a cartel; second, a marker system that allows an applicant to secure its position in line; third, the requirement of full disclosure of evidence; fourth, an ongoing requirement to fully cooperate with the AA; fifth, a requirement to give up collusive behavior. We take these features into account. In line with the US leniency program, in our model only the first evidence-providing firm receives leniency. We require the firms to reveal all evidence in their possession. Furthermore, a reporting firm has to give up its collusive conduct.

Both the US and the EU leniency programs specify requirements concerning the evidence an applicant has to provide. The Antitrust Division of the

[6]Moreover, we embed the ideas of Feess and Walzl (2005) and Silbye (2010) in a more general framework in the spirit of Motta and Polo (2003). In particular, we endogenize the AA's detection and conviction probabilities.

[7]See Department of Justice (2008) and European Commission (2006).

US Department of Justice requires to report "the wrongdoing with candor and completeness" and to provide "cooperation that advances the Division in its investigation."[8] In Section 4, we model these requirements by introducing minimum standards of evidence.

We also use our framework to analyze the effects of a marker system, a standard element of most leniency programs. Common arguments for its use are legal certainty and transparency, and that it encourages a "race to the courthouse." In possession of the marker, a firm has, e.g., 30 days to collect the evidence necessary to "perfect the marker." Within this time, it cannot be leapfrogged by other firms.[9] For reasons of tractability, we ignore the time dimension of the marker system. In our model, the AA immediately informs a firm whether leniency is available.

The rest of this paper is organized as follows. The next section describes the model. In Section 3 we derive the equilibria. In Section 4 we analyze minimum standards of evidence and in Section 5 the marker system. Section 6 concludes.

# 2 Model

We consider a setting where each industry consists of two potentially colluding firms, $a$ and $b$. At the outset, the legislator and the AA announce their policy. The AA's choice consists of two elements. First, the AA starts an investigation with probability $\alpha \in [0,1]$; second, the AA decides how much effort $p \in [0,1]$ it puts into the investigation.[10] This effort gives rise to the probability $P$ of detecting and convicting a cartel; we will specify $P$ as we move along. Moreover, the legislator announces the fine $F > 0$ that

[8] See Department of Justice (2008). Likewise, the European Commission requires that the evidence has to enable the Commission to find an infringement of Article 101 TFEU (European Commission, 2006).

[9] See, e.g., Sandhu (2007).

[10] Actually, the AA determines $\alpha$ and $p$ by the choice of the size and the allocation of its manpower; see Section 2.1. We extend the AA's strategy set in Sections 4 and 5.

a convicted firm pays whenever it communicated with the other firm in the period under consideration. The legislator grants leniency to the first reporting firm.[11] To get leniency, the reporting firm has to provide evidence of the conspiracy and it has to immediately stop the collusive conduct. Without a marker system, the AA does not inform firms about their position in line. If both firms choose to report, nature determines with equal probability who is first. Accordingly, in expectation each firm obtains half the leniency. We look at the case of full leniency so that the reporting firm ends up with no fine while the non-reporting firm pays $F$; if both firms report, each of them pays in expectation $F/2$.

Then an infinitely repeated game starts. The stage game in each period $t = 0, \ldots$ has the following structure: First firm $i$, $i = a, b$, decides whether it wants to communicate with the other firm or not. If both firms choose to communicate, they create evidence that —if detected— leads to a conviction by the AA; unless both firms communicate, they do not engage in illegal behavior and there is thus no evidence thereof. The evidence dissolves at the end of the period.[12]

Then the AA starts an investigation with probability $\alpha$; with probability $(1 - \alpha)$ there is no investigation. Next the firms choose "classical" economic conduct variables such as quantities under Cournot competition, prices under Bertrand competition, or where to sell under exclusive territories. Simultaneously, the firms decide whether they report any communication ($R$) or not ($N$); firms can report both when there is an investigation and when there is no investigation.[13] The firms' reporting behavior together with the AA's

[11] As we will show in the section on minimum standards for evidence, the AA indeed deters cartels at a lower cost with than without a leniency program. Our set-up corresponds to the US system which does not allow for leniency for a second-reporting firm. By contrast, the EU program offers reduced leniency also to all other firms that are not first to come forward, provided that the additional information is sufficiently valuable. Spagnolo (2008), however, argues that the two systems are more similar than they appear.

[12] Note that communication, even when it is not followed by anti-competitive behavior, is considered illegal.

[13] We assume that the firms report whenever they are indifferent. We rule out strategic

investigation effort determines the probability of conviction,

$$P = \begin{cases} p, & \text{if no firm reports;} \\ p_i, & \text{if firm } i \text{ reports, } i = a, b; \\ p_2, & \text{if both firms report.} \end{cases}$$

Let us describe the specific form of $P$ using an example. Suppose the firms' communication has created $m$ pieces of evidence each of which adds $\eta \leq 1/m$ to $P$. Let $m_{AA} \leq m$ denote the evidence uncovered by the AA; $m_{AA} \geq 0$ if the AA has started an investigation and $m_{AA} = 0$ if it did not. $m_i \leq m$ denotes the evidence in possession of firm $i$, $i = a, b$. Then $p = \eta m_{AA}$ is the probability of conviction of the AA's stand-alone evidence; $p$ is chosen by the AA's effort at the outset. Let $\rho_i = \eta m_i$ denote the probability of conviction generated by the stand-alone evidence of firm $i, i = a, b$. Let the pieces of evidence be independently distributed. Then we have for the joint evidence (AA and one firm, or AA and both firms)

$$p_i := 1 - (1-p)(1-\rho_i), \;\; i = a, b;$$
$$p_2 := 1 - (1-p)(1-\rho_a)(1-\rho_b).$$

If, say, only firm $b$ reports, $p_b = p + \rho_b - p\rho_b$; in terms of evidence, expected total evidence is given by the evidence uncovered by the AA plus the evidence provided by firm $b$ minus the expected value of the evidence in joint possession.

Note that we have $p \leq p_i \leq p_2$, $i = a, b$, and subadditivity $p_2 \leq p_a + p_b$. Furthermore, the probability of conviction is equal to one if and only if either the AA has perfect evidence ($p = 1$) or at least one of the firms submits perfect evidence ($\rho_i = 1$, $i = a, b$). Put differently, if $p < 1$ and both firms submit imperfect evidence, the probability of conviction is less than one. Leniency is an instrument for the AA to gather the evidence to

destruction of evidence; see also Footnote 26. If a firm withholds evidence, this will be detected in the course of the investigation and it loses leniency. Accordingly, a firm either submits all its available evidence or it does not report at all.

eventually prove an infringement of the competition law.[14] In our setting, the AA requires that the reporting firm submits all available evidence; this means that firm $i$ reveals $m_i$ pieces of evidence. Nevertheless, it does not require that the reporting firm pleads guilty. The AA still has to prove an infringement.[15] Moreover, the reporting firms do not have to waive their right to appeal the AA's decision.[16] Therefore, we consider the case of a certain conviction when all members of a cartel report an exception rather than the rule.

Without loss of generality we assume $\rho_a \leq \rho_b$, which implies $p_a \leq p_b$. There are various reasons why firms may possess different amounts of evidence. The employees of one firm may be particularly diligent in documenting their communication related to the cartel. A firm complying with, e.g., ISO certification standards has to file the minutes of all meetings while a non-certified firm has no such obligation. Furthermore, a cartel member might have little evidence because an employee involved in the cartel has left the firm in the meantime and all her documents containing information about the cartel have disappeared. Thus, in practice, the overlap of the evidence provided by the firms is typically not perfect.

Next let us describe how firms collude. By communicating firms try to establish collusion with respect to economic conduct.[17] They agree to choose

[14]By contrast, settlement is a tool to speed up the procedure to reach a cartel decision. Both the US Department of Justice and the European Commission require that the parties plead guilty to settle a cartel case.

[15]The US Department of Justice requires that leniency applicants "confess participation in a criminal antitrust violation" (Department of Justice (2008)). However, the Department of Justice has to carry out the investigation and prove an infringement.

[16]For example, in the air cargo cartel Lufthansa received full immunity from fines under the European Commission's leniency program because it was the first to provide information about the cartel (europa.eu/rapid/press-release_IP-10-1487_en.htm?locale=en). Nevertheless, Lufthansa filed an appeal "based on legal considerations" (bloomberg.com/news/2011-01-27/japan-airlines-appeals-48-8-million-antitrust-fine-at-eu-court.html).

[17]In line with Aubert et al. (2006), we assume that tacit collusion is impossible, that is, communication is a prerequisite for collusion.

$q_M$, which is half the monopoly quantity in a Cournot example.[18] This leads to a stage profit of $\pi_M$ for each firm. Firms support the collusive behavior with grim-trigger strategies. If a firm deviates from $q_M$, Nash punishment $q_C$ starts and continues forever; denote the static Nash profit of each firm by $\pi_C$. If a firm deviates while the other firm colludes, the deviating firm chooses $q_D$ and its profit is $\pi_D$; in this case the colluding firm's profit is $\underline{\pi}$. Let $\underline{\pi} \leq \pi_C < \pi_M < \pi_D$. We assume that the cartel is stable in the absence of the AA, meaning that $\pi_M/(1-\delta) \geq \pi_D + \delta\pi_C/(1-\delta)$, where $\delta$ denotes the firms' common discount rate: getting $\pi_M$ forever is better than getting $\pi_D$ in the first round and from then on $\pi_C$.[19] Recall that to get leniency, a reporting firm has to give up its collusive conduct. Therefore, reporting automatically triggers punishment because the reporting firm has to give up setting $q_M$.

## 2.1 Enforcement Technology

Legislation stipulates the fine $F$ as well as full leniency. The AA strives to deter all cartels at minimum cost.[20] It chooses the size of its staff $L$; normalizing the wage to 1, $L$ also measures the AA's cost. Given $N$ industries, the AA determines the $n \leq N$ ones to be investigated, i.e., $\alpha = n/N$. The AA then allocates $L/n$ people to each industry. The stand-alone probability of conviction $p$ increases linearly with the manpower allocated to that industry,

[18] The interpretation of $q_M$ (also of $q_D$ and $q_C$, defined below) depends on the type of the cartel. In a Bertrand set-up, for example, $q_M$ refers to the monopoly price.

[19] In a Cournot set-up with linear demand $1 - q_1 - q_2$ and zero marginal cost, $q_M = 1/4$, $q_C = 1/3$, and $q_D = 3/8$ such that $\underline{\pi} = 7/72$, $\pi_C = 1/9$, $\pi_M = 1/8$, and $\pi_D = 9/64$. The cartel is stable if $\delta \geq 9/17$. In a Bertrand framework firms collude by charging the monopoly price. Deviating means undercutting the monopoly price and competition entails prices equal marginal cost. With exclusive territories firms collude by not selling in their rival's territory. A deviating firm unilaterally sells in the rival's market. If firms compete, they both serve the entire market. See, e.g., Aubert et al (2006).

[20] See Motta and Polo (2003) for a similar approach.

i.e.,

$$p(L,n) = \begin{cases} \gamma L/n, & \text{if } L/n \leq 1/\gamma; \\ 1, & \text{otherwise,} \end{cases}$$

with $\gamma > 0$. If the AA uses manpower efficiently, i.e., $L/n \leq 1/\gamma$, $\alpha p = \gamma L/N$. Therefore, the AA's cost is $L = \alpha p N/\gamma =: C(\alpha p)$. With this enforcement technology it is, for example, possible to monitor all industries ($\alpha = 1$) with very low $p$. This policy implies that the no-investigation subgame is never reached.

# 3 Equilibrium Analysis

We will first derive the firms' behavior in the investigation and the no-investigation subgames. Then we analyze under which conditions firms collude and how the AA can deter communication. Finally, we determine the cost-minimizing deterrence policy as a function of the fine $F$.

## 3.1 The Firms' Behavior

**Investigation Subgame** Let us now analyze the subgame beginning after the firms have communicated and the AA has started its investigation. In our Cournot example, firms choose their quantities. Simultaneously, they decide whether to report or not. As we will see below, their reporting strategies determine their quantity choices. Therefore, we identify the firms' strategies only by their reporting behavior and look at the following equilibrium candidates: Both firms report, $(R, R)$, both firms do not report, $(N, N)$, and one firm reports while the other does not, $(R, N)$ and $(N, R)$. We will next determine under which conditions each candidate can indeed be an equilibrium; if several possible equilibria exist, we pick the Pareto-superior one.

If both firms report, i.e., $(R, R)$, the probability of conviction is $p_2$. The fact that both firms report triggers competition in all future periods. Therefore, their quantity choices are determined solely by the stage game, the only equilibrium of which is the competitive one with both firms setting $q_C$. Accordingly, $(R,R)$ yields $\pi_C/(1-\delta)-p_2F/2$ for each firm. For $(R,R)$ to be an equilibrium, firm $b$, say, must have no incentive to deviate. If $b$ deviates while $a$ reports, the firms still compete from the next period on; thus $b$ sets $q_C$ in the current period. Therefore, if firm $b$ deviates to $N$, it chooses $q_C$ yielding the competitive profit $\pi_C$. Its payoff from deviating is thus $\pi_C/(1-\delta)-p_aF$. Firm $b$ pays the fine for sure rather than with probability 1/2; in return it lowers the probability of conviction from $p_2$ to $p_a$ because the AA has less evidence. Accordingly, a necessary condition for $(R,R)$ to be a Nash equilibrium in the subgame is $p_2 \leq 2p_i,\ i=a,b$. Since $p_a \leq p_b$, $(R,R)$ is an equilibrium if

$$p_2 \leq 2p_a \quad \text{or} \quad p \geq 1-1/((1-\rho_a)(1+\rho_b)) =: \hat{p}. \tag{1}$$

Note that $\hat{p} \leq 1/2$.

If no firm reports, i.e., $(N,N)$, the probability of conviction is $p$. Both firms collude with respect to $q_M$ because this Pareto dominates $(N,N)$ together with $q_C$. Moreover, if it is optimal to collude in the current period, it is also optimal to not report and collude in future periods. Thus, a firm's payoff is $\pi_M - pF + \sum_{t=1}^{\infty}\delta^t(\pi_M-\alpha pF)$.[21] If a firm deviates to $R$ and $q_D$, it earns $\pi_D$ and pays no penalty in the current period. From the next period on the firms compete which implies that there is no risk to pay the fine. The firm's deviation profit is thus $\pi_D+\delta\pi_C/(1-\delta)$. A necessary condition for $(N,N)$ to be a Nash equilibrium in the subgame is

$$p \leq (\pi_M-\delta\pi_C-(1-\delta)\pi_D)/(1-\delta+\delta\alpha)F =: p_{N,N}(\alpha). \tag{2}$$

If $a$ reports while $b$ does not, i.e., $(R,N)$, the probability of conviction is $p_a$. Firm $a$'s reporting triggers competition in all future periods. Thus, the stage game determines the equilibrium and both firms choose $q_C$. Since $a$

[21] This assumes that firms also play $(N,N)$ in the no-investigation subgame. In the next section we show that this is indeed the case.

reports while $b$ does not, $a$ is granted full leniency yielding a payoff $\pi_C/(1-\delta)$. Suppose firm $a$ unilaterally deviates to $N$. Then competition is still triggered from the next period onwards because firm $b$ plays $q_C$ in the current period. Firm $a$ would thus play $N$ together with $q_C$. However, this cannot be optimal because $a$ loses the leniency. Firm $b$ obtains $\pi_C/(1-\delta) - p_a F$ in $(R, N)$. If $b$ also reports, its payoff is $\pi_C/(1-\delta) - p_2 F/2$. Accordingly, a necessary and sufficient condition for $(R, N)$ to be played in the subgame is that the expected fine for $b$ when reporting exceeds the expected fine when not reporting, or

$$p_2 > 2p_a. \tag{3}$$

Analogously, as a condition for $(N, R)$ to be played in the subgame we obtain $p_2 > 2p_b$. The assumption $\rho_a \leq \rho_b$ implies that this condition is never satisfied, so that $(N, R)$ is never played. If (1) is satisfied, (3) does not hold. The two equilibrium candidates $(R, R)$ and $(R, N)$ thus exclude each other. Moreover, one of the candidates always exists.

If (2) is satisfied, the equilibrium $(N, N)$ also exists, so that the issue of equilibrium selection arises. Recall that for $(N, N)$ to be an equilibrium, the equilibrium payoff must be greater or equal than the deviation payoff $\pi_D + \delta\pi_C/(1-\delta)$ which in turn is greater than $\pi_C/(1-\delta)$. In $(R, R)$ both firms get $\pi_C/(1-\delta) - p_2F/2 < \pi_C/(1-\delta)$. In $(R, N)$ firm $a$ gets $\pi_C/(1-\delta)$ and firm $b$ gets $\pi_C/(1-\delta) - p_a F$. Both payoffs are less than the payoffs in $(N, N)$ which, therefore, Pareto-dominates all other possible equilibria. Thus, if (2) is satisfied with strict inequality, the firms indeed play $(N, N)$.

**No-Investigation Subgame** As in the investigation subgame, we can identify the firms' strategies only by their reporting behavior. We will derive the necessary conditions such that firms do not report in the no-investigation case. If they choose to report in this subgame, they will certainly not communicate in the first place so that there is no need for the AA to deter cartel formation.

Our analysis here follows the different cases of the investigation subgame. Let us start with the case where $(R, R)$ is the equilibrium in the investigation subgame, that is, (1) holds and (2) is violated. Suppose firms play $(N, N)$ together with $q_M$ when there is no investigation. Then the ex-ante expected profit from communicating is

$$\pi((N,N),(R,R)) = \sum_{t=0}^{\infty} \delta^t (1-\alpha)^t \left( \alpha \left[ \frac{\pi_C}{1-\delta} - p_2 \frac{F}{2} \right] + (1-\alpha)\pi_M \right).$$

Firms start with communication and continue to do so if there was no investigation in the preceding period; if there is an investigation, they stop to communicate and play $q_C$. Now consider the no-investigation subgame. If a firm does not report, it makes profit $\pi_M + \delta\pi((N,N);(R,R))$. If it reports and chooses $q_D$, its profit is $\pi_D + \delta\pi_C/(1-\delta)$. Accordingly, the firms play $(N, N)$ in the no-investigation subgame if

$$\begin{aligned} &\pi_D + \delta\pi_C/(1-\delta) < \pi_M + \delta\pi((N,N),(R,R)) \quad \text{or} \\ &\pi((N,N),(R,R)) > \pi_C/(1-\delta) + (\pi_D - \pi_M)/\delta. \end{aligned} \tag{4}$$

If (4) is not satisfied, at least one firm reports and picks $q_C$.

Next consider the case where $(R, N)$ is the equilibrium in the investigation case, i.e., (1) and (2) do not hold. Again, we want to determine under which condition both firms do not report in the no-investigation case. Recall that firm $b$ which does not report in the investigation case does worse than the reporting firm which gets leniency. Therefore, if the high-evidence firm $b$ plays $N$ in the no-investigation subgame, the low-evidence firm $a$ will certainly do so, too. Suppose firms play $(N, N)$ together with $q_M$ when there is no investigation. Then the ex-ante expected profit from communicating of firm $b$ is

$$\pi_b((N,N),(R,N)) = \sum_{t=0}^{\infty} \delta^t (1-\alpha)^t \left( \alpha \left[ \frac{\pi_C}{1-\delta} - p_a F \right] + (1-\alpha)\pi_M \right).$$

By the same reasoning as above, the firms play $(N, N)$ in the no-investigation subgame if

$$\pi((N,N),(R,N)) > \pi_C/(1-\delta) + (\pi_D - \pi_M)/\delta. \qquad (5)$$

Finally, consider the case in which $(N, N)$ is the equilibrium of the investigation subgame, i.e., (2) is satisfied. Suppose firms play $(N, N)$ together with $q_M$ when there is no investigation. Then the ex-ante expected profit from communicating is $\pi((N,N),(N,N)) = \sum_{t=0}^{\infty} \delta^t (\alpha [\pi_M - pF] + (1-\alpha)\pi_M)$. The preceding argument yields that the firms do not report in the no-investigation subgame if

$$\pi((N,N),(N,N)) > \pi_C/(1-\delta) + (\pi_D - \pi_M)/\delta.$$

This condition is satisfied if (2) holds.

**Communication Stage and Deterrence** If a firm does not communicate, it obtains $\pi_C$ in the current period. Stationarity implies that the firm does not communicate in all future periods as well. The present value from not communicating is $\pi_C/(1-\delta)$.

Firms communicate and then play $((N,N),(R,R))$ if $\pi((N,N),(R,R)) > \pi_C/(1-\delta)$. The AA rules out this communication equilibrium if it makes sure that the inequality does not hold. Straightforward computations show that this is possible for sufficiently high values of $\alpha$ and $p$. In our set-up, $\alpha$ and $p$ are, however, costly. Obviously it is cheaper to make sure that (4) is not satisfied: then firms report in the no-investigation subgame and this communication equilibrium does not exist. In the appendix we derive the function $p_{R,R}(\alpha)$ such that if $p \geq p_{R,R}(\alpha)$, (4) is not satisfied and this communication equilibrium does not exist.[22]

[22] A proper notation would be $p_{(N,N),(R,R)}(\alpha)$ where $(N, N)$ denotes the firms' strategies in the no-investigation and $(R, R)$ in the investigation subgame. Since for all relevant deterrence constraints firms play $(N, N)$ in the no-investigation subgame, we suppress $(N, N)$ and write as a shortcut $p_{R,R}(\alpha)$.

Firms communicate and play $((N,N),(R,N))$ if $\pi((N,N),(R,N)) > \pi_C/(1-\delta)$. A similar argument as in the previous paragraph shows that the AA deters efficiently by making sure that (5) does not hold. It does so by setting $p \geq p_{R,N}(\alpha)$, which we also derive in the appendix.

Firms communicate and play $((N,N),(N,N))$ if $\pi((N,N),(N,N)) > \pi_C/(1-\delta)$ which holds if $p < (\pi_M - \pi_C)/\alpha F$. This condition is satisfied if (2) holds strictly. Thus, if $p \geq p_{N,N}(\alpha)$, the firms report in the investigation subgame and this communication equilibrium does not exist.

To sum up: The AA achieves complete deterrence if

$$p \geq \max\{p_{N,N}(\alpha), p_{R,R}(\alpha), p_{R,N}(\alpha)\}. \tag{6}$$

## 3.2 Optimal Deterrence

The optimal policy of the AA minimizes $C(\alpha p)$ subject to (6). Proposition 1, which we prove in the appendix, characterizes the solution $(\alpha^*, p^*)$.

**Proposition 1.** *There exist* $0 < \underline{F} < \overline{F}$ *such that:*

*a) if* $F < \underline{F}$, $\alpha^* = (\pi_M - \delta\pi_C - (1-\delta)(\pi_D + F))/\delta F$ *and* $p^* = 1$*;*

*b) if* $F \geq \underline{F}$, $\alpha^* \in (0,1)$ *and* $p^* \in (0,1)$*:*

*i) for* $\hat{p} \leq 0$, $\alpha^*$ *is defined by* $p_{N,N}(\alpha^*) = p_{R,R}(\alpha^*)$*;*

*ii) for* $\hat{p} > 0$ *and* $F \in (\underline{F}, \overline{F}]$, $\alpha^*$ *is defined by* $p_{N,N}(\alpha^*) = p_{R,R}(\alpha^*)$*;*

*iii) for* $\hat{p} > 0$ *and* $F > \overline{F}$, $p_{N,N}(\alpha^*) = p_{R,N}(\alpha^*)$ *defines* $\alpha^*$*.*

To explain this result, first note that the AA must always ensure that the communication equilibrium $((N,N),(N,N))$ is not played by firms. To see this, suppose on the contrary that $p \geq p_{N,N}(\alpha)$ is not binding. If the AA only has to deter the $((N,N),(R,R))$ and $((N,N),(R,N))$ communication equilibria, it can do so by setting, e.g., $\alpha = 1$ and $p$ arbitrarily small, resulting in arbitrarily small enforcement cost. With this policy the no-investigation subgame is never reached. In the investigation subgame firms make the competitive profit $\pi_C$ minus the expected fine. Hence, they do better by not communicating in the first place. Yet, with $p$ small firms prefer not to report in the investigation subgame, i.e., they play $((N, N), (N, N))$.

Part a) of Proposition 1 is illustrated in Figure 1. With small $F$ the legislator endows the AA with limited punishment possibilities. Since fines are low, not reporting in the investigation subgame dominates reporting. Therefore, the AA only has to deter the $((N, N), (N, N))$ communication equilibrium. To do so, high values of $\alpha$ and $p$ are necessary, i.e., $p_{N,N}(\alpha)$ is large. It is cheaper to deter $(N, N)$ in the investigation than in the no-investigation subgame. In the investigation subgame $p$ has a stronger deterrence effect than $\alpha$. Therefore, the AA optimally sets $p^* = 1$ and $\alpha^* < 1$.

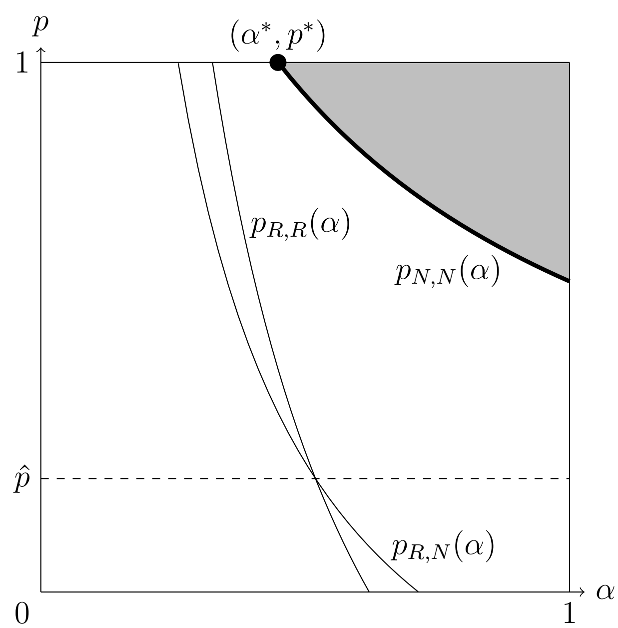


Figure 1: The deterrence region in case a)

Increasing $F$ allows the AA to reduce $\alpha^*$, which leads us to part b) of

Proposition 1. With $\alpha$ small, the no-investigation subgame becomes sufficiently likely, which, in turn, induces firms to collude. The communication equilibria $((N,N),(R,R))$ and $((N,N),(R,N))$ are therefore attractive for firms: they do not report and make profits $\pi_M$ in the no-investigation subgame. Firm $b$ determines which of the two equilibria is played in the investigation subgame. If firm $b$ reports, it increases the probability of conviction by $(p_2 - p_a)$ while it reduces its expected fine by half. Thus, if $p_2 \leq 2p_a$, firm $b$ reports. Recall that $p_2 \leq 2p_a$ is equivalent to $p \geq \hat{p}$ with $\hat{p}$ given by (2); accordingly, if $p < \hat{p}$, firm $b$ prefers not to report and vice versa for $p \geq \hat{p}$. Since $\hat{p} \leq 1/2$, for large values of $p$, firm $b$ reports and $p \geq p_{R,R}(\alpha)$ defines the second binding constraint; see Figure 2. To deter $((N,N),(R,R))$, the AA has to increase $\alpha$; the higher fine $F$ enables the AA to decrease $p$.

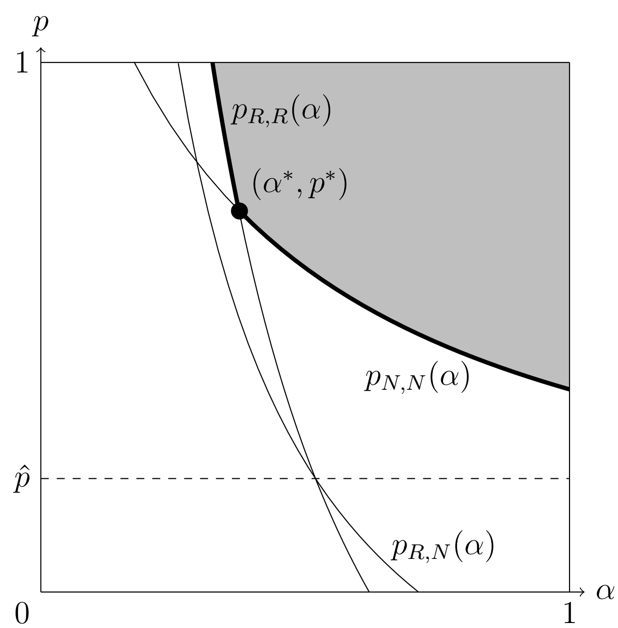


Figure 2: The deterrence region in case b) ii)

Further increasing $F$ allows the AA to lower $p$ until eventually $p < \hat{p}$

(provided $\hat{p} > 0$); see Figure 3. The lower $p$, the higher the increase in the probability of conviction if $b$ reports. Consequently, for low $p$ firm $b$ will not report in the investigation subgame and $p \geq p_{R,N}(\alpha)$ is the second binding constraint. If, however, $\hat{p} \leq 0$, $p \geq p_{R,R}(\alpha)$ continues to define the second binding constraint.

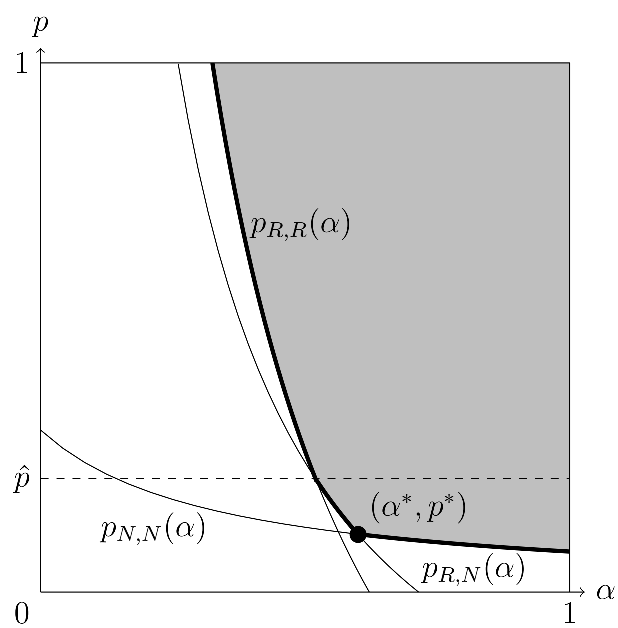


Figure 3: The deterrence region in case b) iii)

The value of $\hat{p}$ thus plays a crucial role as to which constraint is binding. $\hat{p}$ is positive if $\rho_b/(1 + \rho_b) > \rho_a$ which holds for $\rho_b$ sufficiently larger than $\rho_a$. If the two stand-alone probabilities of conviction are of similar size, firm $b$ will report in the investigation subgame like firm $a$. Call this the case of symmetric firms. Let us now make firms asymmetric by increasing $\rho_b$ while holding $\rho_a$ constant. This exercise increases $b$'s expected fine when reporting while $b$'s fine when not reporting is unaffected. Thus, for $\rho_b$ large enough firm $b$ prefers not to report and $p \geq p_{R,N}(\alpha)$ defines the second binding constraint for the AA. $p_{R,N}(\alpha) > p_{R,R}(\alpha)$ for $p < \hat{p}$, meaning that the AA incurs higher cost of deterring cartels with asymmetric firms than cartels with symmetric firms. This case distinguishes our analysis from previous work.[23] These papers look at the case where firms are symmetric and, moreover, reporting of one firm increases the probability of conviction to 1.

To push this point somewhat further: The enforcement cost $C(\alpha^* p^*)$ decreases monotonically in $F$. If the legislator wants to minimize enforcement costs in the spirit of Becker (1968), it chooses a high $F$. If firms are asymmetric, we are in the case of Figure 3. The AA has to deter the firm with the greater evidence from not reporting in the investigation subgame.

# 4 Minimum Standards of Evidence

Let us now analyze a minimum standard of evidence $\hat{\rho}$: a firm is granted leniency if and only if the evidence provided is at least $\hat{\rho}$. If $\hat{\rho} < \rho_a$, the minimum standard has no bite and our preceding analysis applies; $\hat{\rho} > \rho_b$ corresponds to having no leniency at all. As has been shown in the literature, deterrence is more expensive without than with a leniency program: Inducing a firm to report is more difficult because by doing so the firm does not avoid the fine.[24]

This leaves us with the case $\rho_a < \hat{\rho} \leq \rho_b$. Firm $a$ can never avoid the fine so that firm $b$ is the only one to obtain leniency. Accordingly, the communication equilibria $((N, N), (R, R))$ and $((N, N), (R, N))$ no longer exist. To deter the $((N, N), (N, N))$ communication equilibrium, the AA efficiently induces firm $b$ to report in the investigation subgame because it qualifies for leniency. Firm $b$'s payoffs from not reporting and deviating are as described above. Therefore, the AA deters this communication equilibrium by setting

[23] See, e.g., Motta and Polo (2002), Aubert et al (2006), and Harrington (2008).

[24] See, e.g., Motta and Polo (2002), Aubert et al (2006), and Harrington (2008). Formally, the $p_{N,N}(\alpha)$-curve which always binds in the optimum shifts upwards without leniency. The lowest cost of deterrence along this curve exceeds the highest cost of deterrence along the $p_{N,N}(\alpha)$-curve when leniency is available.

$p \geq p_{N,N}(\alpha)$, with $p_{N,N}(\alpha)$ defined by (2).

Next consider $((N,N),(N,R))$. In the no-investigation subgame, both firms have the same deviation profit, $\pi_D + \delta\pi_C/(1-\delta)$. Firm $a$, however, has the smaller continuation profit since it does not qualify for leniency. Therefore, the AA optimally induces firm $a$ to deviate. In the no-investigation subgame, firm $a$'s payoff from colluding is $\pi_M + \sum_{t=1}^{\infty} \delta^t (1-\alpha)^{t-1} \left[\alpha\left(\pi_C/(1-\delta) - p_b(p)F\right) + (1-\alpha)\pi_M\right]$. Firm $a$ deviates if

$$p \geq \frac{\pi_M - \pi_D + \delta(1-\alpha)(\pi_D - \pi_C) - \rho_b \delta\alpha F}{(1-\rho_b)\delta\alpha F} =: p_{N,R}(\alpha).$$

Note that $p_{N,R}(\alpha)$ equals $p_{R,N}(\alpha)$, with $\rho_b$ substituted by $\rho_a$. Accordingly, $p_{N,R}(\alpha)$ qualitatively has the same properties as $p_{R,N}(\alpha)$.

Like in Proposition 1, for $F \geq \underline{F}$ the AA's optimal choice $(\alpha^*, p^*)$ satisfies $p_{N,N}(\alpha^*) = p_{N,R}(\alpha^*)$, i.e., both restrictions hold with equality. Straightforward computations show that $p \geq p_{N,R}(\alpha)$ is less stringent than both $p \geq p_{R,R}(\alpha)$ and $p \geq p_{R,N}(\alpha)$. Consequently, $C(\alpha^* p^*)$ is lower with the minimum standard than without. Given that firm $a$ never applies for leniency, firm $b$ enjoys immunity whenever it reports. If firm $b$ reports, the AA gets ample of evidence. This, in turn, implies that firm $a$ ex ante faces a high expected fine. By contrast, without minimum standards either both firms pay $F/2$ when convicted along $((N,N),(R,R)))$, or the conviction rate is $p_a < p_b$ because only the low-evidence firm $a$ reports along $((N,N),(R,N))$.

To sum up: A minimum standard may reduce the AA's enforcement cost if firms are asymmetric and the standard is such that firm $b$ qualifies for leniency while firm $a$ does not. Such a minimum standard provides the high-evidence firm with strong incentives to report: it can be sure to avoid the fine if it chooses to report; the low-evidence firm cannot interfere. Nevertheless, a minimum standard requires that the evidence is quantifiable and verifiable. Moreover, the AA needs to know how much evidence the firms have to get the standard right.[25] If the AA faces, e.g., uncertainty as to the evidence

[25]In our model $\rho_a$ and $\rho_b$ are common knowledge.

possessed by firms, it runs the risk of setting the standard too high so that it actually does worse than with no minimum standard at all.[26]

# 5 Marker System

## 5.1 Conditional Reporting

Under a marker system the AA informs a firm upon request about its position in line concerning leniency. If the firm is the first one in line, it gets the marker and leniency if it chooses to report. By contrast, if the firm is second in line and the first firm reports, the AA informs the second firm that the marker and thus leniency are no longer available. Should both firms request information about the availability of the marker at the same time, nature determines with equal probability who is first.

The marker system creates additional strategies for firms compared to the no-marker case. They can request the information and then condition their reporting behavior on whether they get the marker ($m$) or not ($n$). For example, a firm can play the strategy: report if the marker is available and do not report if the marker is not available, i.e., $(R|m, N|n)$. This strategy weakly dominates unconditional reporting $R$. If, say, firm $a$ is the first one to report, both strategies yield the same payoff. If, however, firm $a$ is second in line, $R$ yields a payoff strictly lower than $(R|m, N|n)$. Since the firm does not get the marker, it has to pay the fine $F$ if convicted under both strategies. If it does not report, the probability of conviction is lower than if it reports. Likewise, the strategy $(N|m, R|n)$ is weakly dominated by unconditional not reporting $N$. If firm $a$ is first in line, both strategies yield the same payoff. If, however, firm $a$ is second in line, $(N|m, R|n)$ yields a payoff strictly lower

[26] Note that with minimum standards firms have an incentive to keep evidence in order to qualify for leniency (see also Aubert et al. (2006) and Agislaou (2012)). Therefore, our assumption that evidence cannot be destroyed is less restrictive with minimum standards than without. This may be seen as a further argument in favor of minimum standards of evidence.

than $N$. The firm has to pay the fine $F$ if convicted under both strategies. If it does not report, the probability of conviction is lower than if it reports.

Assuming that firms do not play weakly dominated strategies, we are left with the two strategies $(R|m, N|n)$ and $N$ for both the investigation and the no-investigation subgame. As a shortcut we will denote the conditional strategy $(R|m, N|n)$ by $RN$. The two communication equilibria $((N, N), (N, N))$ and $((N, N), (RN, RN))$ are thus of interest. The analysis of the $((N, N), (N, N))$ equilibrium is along the same lines as in the no-marker case. The deviation to consider is $RN$ rather than $R$. When first in line, the deviation to $RN$ generates the same payoffs as the deviation to $R$. Thus, the AA deters this equilibrium if $p \geq p_{N,N}(\alpha)$, where $p_{N,N}(\alpha)$ is defined in (2). Let us now turn to the $((N, N), (RN, RN))$ communication equilibrium.

## 5.2 The Firms' Behavior

**Investigation Subgame** If both firms play $RN$, the probability of conviction is with equal probability $p_a$ or $p_b$, depending on who is first in line. Firm $b$, for example, gets a payoff $\pi_C/(1-\delta) - p_a F/2$; if $b$ deviates to $N$, its payoff amounts to $\pi_C/(1-\delta) - p_a F$. Consequently, $(RN, RN)$ is always an equilibrium in the investigation subgame. If $(N, N)$ is an equilibrium in the investigation subgame, it Pareto dominates $(RN, RN)$ and firms play $(N, N)$. This is the case if $p < p_{N,N}(\alpha)$.

**No-Investigation Subgame** Suppose firms play $(N, N)$ together with $q_M$ when there is no investigation. Then firm $a$'s ex-ante expected profit from communicating is

$$\pi_a((N, N), (RN, RN)) = \sum_{t=0}^{\infty} \delta^t (1-\alpha)^t \left( \alpha \left[ \frac{\pi_C}{1-\delta} - p_b \frac{F}{2} \right] + (1-\alpha)\pi_M \right).$$

Note that $b$'s profit is higher than $a$'s because $p_a < p_b$. Now consider the no-investigation subgame. If firm $a$ does not report, it makes profit

$\pi_M + \delta\pi_a((N,N),(RN,RN))$. If it reports and chooses $q_D$, its profit is $\pi_D + \delta\pi_C/(1-\delta)$. Accordingly, $(N,N)$ is an equilibrium of the no-investigation subgame if

$$\pi_a((N,N),(RN,RN)) \geq \pi_C/(1-\delta) + (\pi_D - \pi_M)/\delta. \tag{7}$$

If (7) is not satisfied, at least one firm reports and picks $q_C$.

**Communication Stage and Deterrence** To deter $((N,N),(N,N))$ the AA sets $p \geq p_{N,N}(\alpha)$. To deter $((N,N),(RN,RN))$ the AA has to make sure that (7) is not satisfied: then firms report in the no-investigation subgame and this communication equilibrium does not exist. If $p \geq (2(\pi_M - \pi_D) + 2\delta(1-\alpha)(\pi_D - \pi_C) - \rho_b\delta F)/(1-\rho_b)\delta\alpha F =: p_{RN,RN}(\alpha)$, (7) is not satisfied and this communication equilibrium does not exist.

To sum up the marker case: The AA achieves complete deterrence if

$$p \geq \max\{p_{N,N}(\alpha), p_{RN,RN}(\alpha)\}. \tag{8}$$

## 5.3 Optimal Deterrence

The optimal policy of the AA minimizes $C(\alpha p)$ subject to (8). Proposition 2 characterizes the solution $(\alpha^*, p^*)$.[27]

**Proposition 2.** *There exists* $\underline{F} > 0$ *such that:*

*a) if* $F < \underline{F}$, $\alpha^* = (\pi_M - \delta\pi_C - (1-\delta)(\pi_D + F))/\delta F$ *and* $p^* = 1$*;*

*b) if* $F \geq \underline{F}$, $\alpha^* \in (0,1)$ *and* $p^* \in (0,1)$ *with* $\alpha^*$ *defined by* $p_{N,N}(\alpha^*) = p_{RN,RN}(\alpha^*)$.

In Figure 4, the bold dashed line depicts the set of the optimal deterrence combinations $(\alpha^*, p^*)$; for reasons of clarity we have not included the $p_{N,N}$-curve. Let us now compare the enforcement cost under the no-marker and

[27] Since the proof of Proposition 2 is similar to the proof of Proposition 1, we skip it.

the marker system. If $F < \underline{F}$ so that $p^* = 1$, the AA deters with the same $(\alpha^*, p^*)$ combinations under both systems. Accordingly, the AA's deterrence costs are the same.

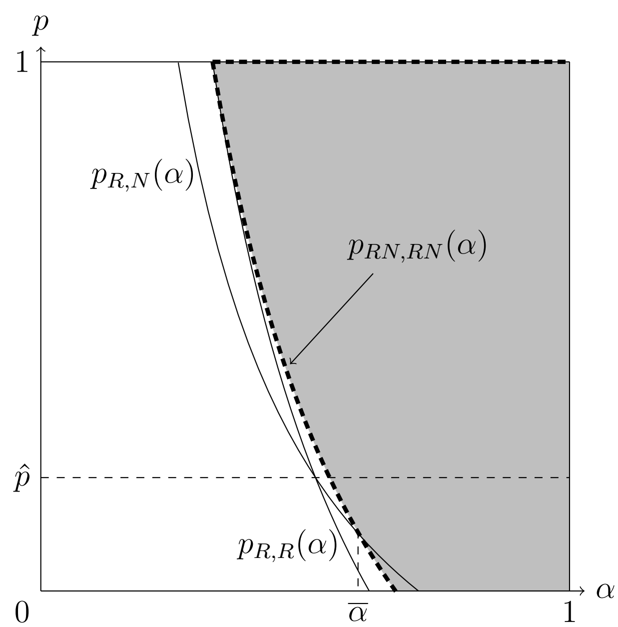


Figure 4: The deterrence region in the marker scenario

If $F > \underline{F}$, $p^* = p_{RN,RN}(\alpha^*) < 1$. Furthermore, $p_{R,R}(\alpha) < p_{RN,RN}(\alpha)$ whenever the latter is less than 1. This means that for the cases b) i) and ii) of Proposition 1 deterrence is more expensive in the marker than in the no-marker scenario. In the no-marker scenario the AA deters the $((N,N),(R,R))$ equilibrium where in the investigation subgame the probability of conviction is $p_2$. In the marker scenario the AA deters the $((N,N),$ $(RN,RN))$ equilibrium where in the investigation subgame the expected probability of conviction is $(p_a + p_b)/2 < p_2$. The expected payoff from colluding is thus higher in the marker than in the no-marker case and a higher $p$ is necessary to deter firms from communicating.

This leaves us with the case b) iii) of Proposition 1. Here the outcome depends on whether $p_{RN,RN}(\alpha)$ and $p_{R,N}(\alpha)$ intersect at some $\bar{\alpha} > 0$. If they do not intersect, a marker is detrimental. Yet, if they intersect, for $\alpha^* > \bar{\alpha}$, $p_{RN,RN}(\alpha^*) < p_{R,N}(\alpha^*)$, and a marker is beneficial; see Figure 4. In the $((N,N),(R,N))$ equilibrium, the probability of conviction is $p_a$ in the investigation subgame. In the $((N,N),(RN,RN))$ equilibrium, the probability of conviction is $(p_a + p_b)/2 > p_a$ in the investigation subgame. The expected payoff is thus higher in the no-marker case than in the marker case and a higher $p$ is necessary to deter firms from colluding.

To sum up: Deterrence is cheaper under the marker than the no-marker system if and only if $\hat{p}$ and $F$ are sufficiently large. $\hat{p}$ is large when firms are sufficiently asymmetric, i.e., $\rho_b \gg \rho_a$. In the no-marker case only firm $a$ reports whereas in the marker case both firms report with equal probability. Colluding is thus less attractive in the marker case and the AA has an easier time to deter firms from communicating. In all other cases the marker system does worse because the second-in-line firm withholds its evidence.

Finally, let us look at a minimum standard of evidence $\hat{\rho}$ under the marker system. If $\hat{\rho} < \rho_a$, the minimum standard has no bite and the preceding analysis applies; $\hat{\rho} > \rho_b$ corresponds to no leniency. If $\hat{\rho} \in (\rho_a, \rho_b]$, firm $b$ is the only one to be able to obtain leniency; firm $a$ cannot interfere. Firm $b$ does not need a marker to forestall being queue-jumped by firm $a$; therefore, our analysis from Section 4 applies.

# 6 Conclusions

Nearly the entire literature on leniency looks at the case where all members of a cartel have perfect evidence of the illegal conduct. This assumption implies that if one member reports, no matter which one, the cartel is convicted for sure. If all firms have perfect evidence, the distribution of evidence is symmetric and either all firms apply for leniency or no firm at all, a feature which is typically not observed in practice.[28] Moreover, within such a framework one cannot meaningfully analyze minimum-evidence standards and marker systems, two typical features of actual leniency programs.

Our analysis highlights the importance of appropriate minimum standards of evidence. In a setting with high fines and asymmetric evidence, deterrence is costly since the high-evidence firm has no incentive to report. The AA can lower its enforcement cost by choosing a minimum standard of evidence such that the high-evidence firm qualifies for leniency while the low-evidence firm does not. Such a minimum standard restores the incentives to report for the high-evidence firm. It can be sure to avoid the fine, because the low-evidence firm is, in fact, excluded from leniency. The challenge for the AA in practice is to get the minimum standard right. If the standard is so low that both firms fulfill the conditions for leniency, it remains ineffective. By contrast, a standard so high that no firm reaches the threshold renders the whole leniency program ineffective.

When minimum evidence standards can be used, there is no need for a marker system: The high-evidence firm is protected from being queue-jumped anyway. If minimum evidence standards are not feasible, a marker system has an impact. With a marker the AA gets only one report, with equal probability the high- or the low-evidence one. This increases the deterrence cost compared to the no-marker set-up if without the marker both firms report; this is the case if the firms have similar evidence. If, however, without the marker only the low-evidence firm turns in, the marker lowers the AA's deterrence cost; this happens if firms possess sufficiently different evidence. Our model, therefore, suggests to use a marker system if and only if the AA expects asymmetric distributions of the evidence in the possession of the colluding firms.

We have not dealt with the arguments made in favor of a marker system such as that this institution creates legal certainty and transparency or that

[28]See, e.g., Harrington (2013).

the possibility to secure a marker with preliminary evidence may induce a race to the courthouse. The analysis of these arguments to justify marker systems is left for future research.

# Appendix

Properties of $p_{R,R}$ and $p_{R,N}$:

Using $p_2(p) = 1-(1-p)(1-\rho_a)(1-\rho_b)$, write (4) as

$$p \geq \frac{2(\pi_M-\pi_D)+2\delta(1-\alpha)(\pi_D-\pi_C)-(\rho_a+\rho_b-\rho_a\rho_b)\delta\alpha F}{(1-\rho_a)(1-\rho_b)\delta\alpha F} =: p_{R,R}(\alpha). \quad \text{(A1)}$$

$p_{R,R}(\alpha)$ is continuous and decreasing in $\alpha$. Rewriting (A1) yields

$$\delta\alpha F\left[p(1-\rho_a)(1-\rho_b)+(\rho_a+\rho_b-\rho_a\rho_b)\right] \geq 2(\pi_M-\pi_D)+2\delta(1-\alpha)(\pi_D-\pi_C). \quad \text{(A2)}$$

The LHS of (A2) equals 0 for $\alpha = 0$ and increases in $\alpha$. The RHS is positive for $\alpha = 0$ because $\pi_M/(1-\delta) \geq \pi_D + \delta\pi_C/(1-\delta)$ and negative for $\alpha = 1$. Hence, (4) holds for values of $\alpha$ below some threshold and does not hold for values of $\alpha$ above this threshold.

Using $p_a(p) = 1-(1-p)(1-\rho_a)$, write (5) as

$$p \geq \frac{\pi_M-\pi_D+\delta(1-\alpha)(\pi_D-\pi_C)-\rho_a\delta\alpha F}{(1-\rho_a)\delta\alpha F} =: p_{R,N}(\alpha). \quad \text{(A3)}$$

$p_{R,N}(\alpha)$ is continuous and decreases in $\alpha$ for $p_{R,N}(\alpha) \geq 0$. Rewrite (A3) as

$$\delta\alpha F\left[p(1-\rho_a)+\rho_a\right] \geq \pi_M-\pi_D+\delta(1-\alpha)(\pi_D-\pi_C). \quad \text{(A4)}$$

(A4) holds for values of $\alpha$ close to 1 and is violated for values of $\alpha$ small. Hence, (5) holds for values of $\alpha$ below some threshold and does not hold for values of $\alpha$ above this threshold.

Proof of Proposition 1:

Rewrite the AA's minimization problem as $(\alpha^*, p^*) = \arg\min_{\alpha,p} C(\alpha p)$

$$\text{s.t. } p \geq p_{N,N}(\alpha) \quad \text{or} \quad \alpha \geq \alpha_{N,N}(p), \quad \text{(A5)}$$

$$p \geq p_{R,R}(\alpha) \quad \text{or} \quad \alpha \geq \alpha_{R,R}(p), \quad \text{(A6)}$$

$$p \geq p_{R,N}(\alpha) \quad \text{or} \quad \alpha \geq \alpha_{R,N}(p), \quad \text{(A7)}$$

$$0 \leq \alpha \leq 1, \quad \text{and} \quad 0 \leq p \leq 1,$$

where

$$\alpha_{N,N}(p) := \frac{\pi_M - \delta\pi_C - (1-\delta)\pi_D}{\delta p F} - \frac{1-\delta}{\delta},$$

$$\alpha_{R,R}(p) := \frac{\pi_M - \delta\pi_C - (1-\delta)\pi_D}{\delta\left(\left[p(1-\rho_a)(1-\rho_b) + \rho_a + \rho_b - \rho_a\rho_b\right]F/2 + \pi_D - \pi_C\right)},$$

$$\alpha_{R,N}(p) := \frac{\pi_M - \delta\pi_C - (1-\delta)\pi_D}{\delta\left(\left[p(1-\rho_a) + \rho_a\right]F + \pi_D - \pi_C\right)}.$$

First note that $p^* \geq p_{N,N}(1) > 0$ because $p_{N,N}$ is decreasing in $\alpha$; $\alpha^* \geq \alpha_{R,R}(1) > 0$ because $\alpha_{R,R}$ is decreasing in $p$.

Second note that (A5) is always binding. Consider the relaxed problem of minimizing the cost subject to (A6) and (A7). Then $\alpha = 1$, $p = \epsilon > 0$ and small satisfy (A6) and (A7). The no-investigation subgame is never reached. In the investigation subgame firms always report. They earn $\pi_C$ and pay the fine with positive probability $\epsilon$. They do better by not reporting in the first place. This policy gives rise to the cost $C(1\epsilon) = \epsilon L/\gamma$. Obviously, $C(\alpha^* p^*) \leq \epsilon L/\gamma$. Now suppose (A5) is slack. Using $\alpha^* \geq \alpha_{R,R}(1) > 0$ we have

$$\alpha^* p^* \geq \frac{\pi_M - \delta\pi_C - (1-\delta)\pi_D}{F\left((1-\delta)/\alpha^* + \delta\right)} > \frac{\pi_M - \delta\pi_C - (1-\delta)\pi_D}{F\left((1-\delta)/\alpha_{R,R}(1) + \delta\right)} > 0,$$

meaning $\alpha^* p^* > \epsilon L/\gamma$ for $\epsilon$ sufficiently small, contradicting $C(\alpha^* p^*) \leq \epsilon L/\gamma$.

Third note that $(\alpha^*, p^*) = (\hat{\alpha}, 1)$ with $\hat{\alpha} = \left(\pi_M - \delta\pi_C - (1-\delta)(\pi_D + F)\right)/\delta F$ if and only if (A6) and (A7) are not binding. Consider the Lagrangian of minimizing $C$ subject to (A5). Solving the first-order conditions with respect to $\alpha$ and $p$ yields $1-\delta = 0$, a contradiction. Hence, either $\alpha^* \in \{0,1\}$ or $p^* \in \{0,1\}$. We know already that $\alpha^* > 0$ and $p^* > 0$. Consequently, either $\alpha^*$ or $p^*$ equals 1. Suppose $\alpha^* = 1$ and $p^* = (\pi_M - \delta\pi_C - (1-\delta)\pi_D)/F$. Straightforward computations show that decreasing $\alpha$ by $d\alpha$ and increasing $p$ by $d\alpha(1-\epsilon)p/\alpha$ lowers the cost without violating (A5). Therefore, $p^* = 1$ and $\alpha^* = \hat{\alpha}$.

Fourth note that $(\alpha^*, p^*) = (\hat{\alpha}, 1)$, or equivalently (A6) and (A7) do not hold, if and only if $F \leq \underline{F}$. Since $\alpha_{R,R}(1) \geq \alpha_{R,N}(1)$, (A6) and (A7) are not binding at $(1, \hat{\alpha})$ if

$$2\left(\pi_M - \delta\pi_C - (1-\delta)\pi_D\right)(\pi_D - \pi_C) \geq \left(\pi_M - \pi_C + (1-\delta)(\pi_D - \pi_C)\right)F + (1-\delta)F^2.$$

The LHS is positive. The RHS is 0 for $F = 0$ and strictly increasing in $F$. Thus, there exists a unique $\underline{F}$ with the desired properties.

Fifth note that for $F > \underline{F}$, $\alpha^* \in (0,1)$ and $p^* \in (0,1)$. We know from the previous step that for $F > \underline{F}$ either (A6) or (A7) is binding. Since $\alpha_{R,N}(p) \in (0,1)$ and $\alpha_{R,R}(p) \in (0,1)$ for all $p \in [0,1]$, $\alpha^* \in (0,1)$. From our first step we know that $p^* > 0$. From our third step we know that in this case $\alpha^* = \hat{\alpha}$. In this case, however, if $p^* = 1$ step 4 implies $F < \underline{F}$, a contradiction.

Sixth, straightforward computations show that $\alpha_{R,R}(p) \geq \alpha_{R,N}(p)$ or equivalently $p_{R,R}(\alpha) \geq p_{R,N}(\alpha)$ if and only if $p \geq \hat{p}$ with $\hat{p}$ defined in (1).

Seventh, if $F > \underline{F}$ and $\hat{p} \leq 0$, $p^* = p_{N,N}(\alpha^*) = p_{R,R}(\alpha^*)$. From step 2 we know that $p^* = p_{N,N}(\alpha^*)$ and steps 3 and 4 imply $p^* = \max\{p_{R,R}(\alpha^*), p_{R,N}(\alpha^*)\}$. By step 6, $p^* = p_{R,R}(\alpha^*)$ if and only if $p^* \geq \hat{p}$. Since $\hat{p} \leq 0$ and $p^* > 0$, this is always true.

Eighth, If $\hat{p} > 0$, $p^* = p_{N,N}(\alpha^*) = p_{R,R}(\alpha^*)$ for $F \in [\underline{F}, \overline{F}]$ and $p^* = p_{N,N}(\alpha^*) = p_{R,N}(\alpha^*)$ for $F \geq \overline{F}$. For $F > \underline{F}$ we have $p^* = p_{N,N}(\alpha^*) = \max\{p_{R,R}(\alpha^*), p_{R,N}(\alpha^*)\}$. As can be easily shown, the three functions have unique intersections. Therefore, $p^*$ is unique and, moreover, continuous and decreasing in $F$. For $F = \underline{F}$, $p^* = 1$. If we increase $F$ slightly, by continuity $p^*$ falls slightly. Hence, $p^* > 1/2 \geq \hat{p}$ and by step 6 $p_{R,R} > p_{R,N}$. For $F$ sufficiently large, $p^*$ is arbitrarily small, meaning $p^* < \hat{p}$. Consequently, $p_{N,R} > p_{R,R}$. ■

# References


Agisilaou, P. 2012. Keep to sustain or keep to exploit? Why firms keep hard evidence, Munich CCP Working Paper 12-5, mpra.ub.uni-muenchen.de/39109/

Aubert, C., Kovacic, W. and P. Rey, 2006. The Impact of Leniency Programs on Cartels, *International Journal of Industrial Organization* 6, 1241-1266.

Becker, G., 1968. Crime and Punishment: An Economic Approach, *Journal of Political Economy* 76, 169–217.

Bigoni, M., Fridolfsson, S., Le Coq, C. and G. Spagnolo, 2012. Fines, Leniency, and Rewards in Antitrust, *Rand Journal of Economics* 43, 368–390.

Brenner, S. 2009. An Empirical Study of the European Corporate Leniency Program, *International Journal of Industrial Organization* 27, 639-645.

Department of Justice, 2008. Frequently Asked Questions Regarding the Antitrust Division's Leniency Program and Model Leniency Letters, justice.gov/atr/public/criminal/239583.htm.

European Commission, 2006. Commission Notice on Immunity from Fines and Reduction of Fines in Cartel Cases, ec.europe.eu/competition/ cartels/legislation/leniency_legislation.html.

Harrington, J., 2008. Optimal Corporate Leniency Programs, *Journal of Industrial Economics* 56, 215-246.

Harrington, J., 2013. Corporate Leniency Programs when Firms have Private Information: The Push of Prosecution and the Pull of Pre-Emption, *Journal of Industrial Economics* 61, 1-27.

Harrington, J. and M. Chang, 2009. Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy, *Journal of the European Economic Association* 7, 1400-1435.

Feess, E. and M. Walzl, 2005. Quid-Pro-Quo or Winner-Takes-It-All? An Analysis of Corporate Leniency Programs and Lessons to Learn for US and EU Policies, arno.unimaas.nl/show.cgi?fid=3978.

Miller, N., 2009. Strategic Leniency and Cartel Enforcement,*American Economic Review* 99, 750-768.

Motta, M. and M. Polo, 2003. Leniency Programs and Cartel Prosecution, *International Journal of Industrial Organization* 21, 347-379.

Rey, P., 2003. Towards a Theory of Competition Policy, in Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, M. Dewatripont, L.P. Hansen, S.J. Turnovsky (eds.), 2003 Cambridge University Press, 82–132.

Sandhu, J., 2007. The European Commission's Leniency Policy: A Success?, *European Competition Law Review* 28, 148–157.

Sauvagnat, J., 2014. Prosecution and Leniency Programs: The Role of Bluffing in Opening Investigations, *Journal of Industrial Economics* forthcoming.

Silbye, F., 2010. Asymmetric Evidence and Optimal Leniency Programs in, Topics in Competition Policy: Cartels, Leniency, and Price Discrimination, PhD Thesis, University of Copenhagen, Copenhagen, 41–61.

Spagnolo, G., 2003. Divide et Impera: Optimal Deterrence Mechanisms Against Cartels and Organized Crime, sites.google.com/site/giancarlospagnoloshomepage/research-2/working-papers.

Spagnolo, G., 2008. Leniency and Whistleblowers in Antitrust, in Paolo Buccirossi (ed.), Handbook of Antitrust Economics, MIT Press, Cambridge, Massachusetts, USA, 259–303.